AIPL INFOSHEET
DP2 (9-97)

AIPL Data Processing Update for DRPC Workshop (September 1997)

L.M. Walton
Animal Improvement Programs Laboratory, ARS, USDA, Beltsville, MD 20705-2350
301-504-8334 (voice) • 301-504-8092 (fax) • rlaipl@ggpl.arsusda.gov • http://aipl.arsusda.gov

Format Changes

Participants at the April 1997 DRPC Workshop agreed that the lactation error record (**format 4T•E**) should contain the 710 bytes of format 4T, a disposition code, and 6 error segments. The error segments in the new system are longer to accommodate the 17-byte identification (ID) number and 4-digit year. Likewise, the pedigree error record (**format 1N•E**) for returning error data to organizations that provide pedigrees contains format-1N fields, a disposition code, and 6 error segments.

A replacement for herd-average format 14 (**format 14N**) is proposed. Format 14N has a 2-character breed code and will be used to verify information on herd testing characteristics reported in format 4T. A detailed handout on format 14N indicates the location of equivalent fields in the current format 14. Where no equivalent field is indicated, additional information is being requested.

The uniform type formats also have been modified to accommodate international ID numbers and a 4-digit year. Format numbers have been changed as follows. Format 700, which is used by breed associations to report type scores to the Animal Improvement Programs Laboratory (AIPL), is now **format 7**. Format 940S, which is used by AIPL for reporting bull evaluations for type, is now **format 40**. Format 941, which is used by AIPL for reporting cow evaluations for type, is now **format 41**.

Bull evaluation **format 38** was presented for comment at the April 1997 DRPC Workshop. Because no feedback was received, AIPL considers this format to be final. **Format 380S** will be retired after the release of November 1997 USDA-DHIA genetic evaluations. An evaluation file with the same selection criteria will be provided for future releases using format 38.

Format 105N (cow evaluations) has been used since February 1997. Extra bytes were reserved for the initiation of international ID. Beginning with the release of February 1998 evaluations, animal ID fields will contain true international ID numbers (2-character breed code, 3-character country code), and Canadian ID numbers will not contain the 04 prefix.

Data exchange for February 1998 evaluations

To accommodate the new international ID system and calendar dates for the next millennium and to facilitate movement from a minicomputer to a workstation environment, the current AIPL editing system is being rewritten using a commercial database management package for workstations. The conversion process is expected to continue through the end of 1997. Because the new editing system will not be functional immediately after the record cutoff date for November 1997 evaluations, AIPL has developed the following interim data exchange procedure for February 1998 evaluations.

After the November cutoff, lactation and pedigree data should be submitted in formats 4T and 1N, respectively, which AIPL will convert to formats 4 and 1 during the interim before implementation of the new editing system. To assist AIPL in the conversion process, some small revisions have been made to format 4T. Fields for the number of supervised tests (bytes 232-233) and the number of tests for which components were taken (bytes 234-235) were added to format 4T so that the presence of all test-day segments is not necessary for processing of records. These new fields are especially important for cows that change herds and, therefore, should not have tests in the previous herd included in the record.

The delete options for lactation-type code (position 126) have been clarified. Converted records are validated with the present editing system, and any resulting error records are formatted as 4T•E or 1N•E if requested, otherwise they will be provided in format 4E or 1E.

The current editing system can process only records with 10-character ID information (1-character breed code and 9-character ID number). For interim processing, international ID in format-4T and 1N records is converted by AIPL to 10-character ID if possible. If conversion is not possible, records that include international ID are saved until the new editing system is operational (possibly for February 1998 evaluations but definitely for May 1998 evaluations).

In summary, data exchange for February 1998 evaluations will use formats 4T and 1N for incoming lactation and pedigree records and formats 4T•E and 1N•E for any resulting error records. Records with international ID that

cannot be represented as 10-character ID will not be processed until the new editing system is functional.

Future data exchange procedures (electronic)

The following discussion focuses on the exchange of lactation data, but the same file-naming techniques, frequency of submission, and processing methods apply to pedigree data submitted by breed associations.

To balance the workload associated with data exchange and to provide all data providers with quick turnaround, AIPL would like to receive lactation data on a daily basis. With daily receipt of data, electronic data exchange is a necessity. The proposed protocol for data exchange follows.

All herd information processed by a DRPC during a day plus any correction records are placed in a file (or files) in the order in which they should be processed (AIPL no longer will sort data), compressed using PKZIP-compatible software, and then transferred to the AIPL file transfer protocol (FTP) server (**ftpaipl.arsusda.gov**). The FTP server requires a specific user identification and a password, and AIPL should be contacted to activate each account. With daily submission of data, usually only one test-day segment is needed. To conserve disk space, place an end-of-record indicator after the last test-day segment of the format-4T record. To provide for multiple uniquely named compressed files per day, the naming convention must be an 8-character date with a suffix of ‘z‘ or ‘z00' thru ‘z99' (e.g., 19970925.z01). Format 4T files within a compressed file should adhere to the naming convention of an 8-character date with a suffix of **4T** (e.g., 19970925.4T). Because much of the data exchanged is set up for personal computers, a file-naming convention was developed that adheres to typical DOS “8+3” file naming. By using a number or character in the last position of the **4T** suffix, many files can be uniquely identified in one compressed file. This last character also is used to determine the order in which files are processed. A file with no additional character will be processed first and followed by files with **0** through **9** and then files with **a** through **z**.

All data files with names adhering to the naming convention above are processed via the editing system overnight. Each file is processed separately and in the order indicated by the last character of the input file name. Any error file resulting from the editing process has the same name as the input file with the format letter portion of the suffix changed to **E** (e.g., error file 19970925.4E0 would result from editing data file 19970925.4T0). Files are compressed using PKZIP-compatible software and placed in the **/out** directory of a DRPC’s account on the AIPL FTP server. All error files generated for a given day are placed in one compressed file, and the 8-character creation date followed by a **zip** suffix is the file name (e.g., 19970925.zip).

During the interim period following November 1997 evaluations and prior to implementation of the new editing system, data are accepted on a daily basis for February 1997 evaluations. However, because the design of the current editing system makes daily processing of files quite difficult, daily records are collected for 2 weeks and processed in a group.

Online access to error records

Online access to error records is being developed. Record selection is based on herd, herd-control number, ID number, sire, and artificial-insemination (AI) organization. Only the latest error record for a calving date is retained. Selection for sire and AI organization returns an error record only if no usable record is found for the cow's lactation and the lactation number is 5 or less. After release of evaluations, records submitted more than 6 months prior are deleted. Format-1 errors are in the same database table as format-4 errors; therefore, a query on animal ID number returns errors from both sources. This online application should be available after the release of February 1998 evaluations. The table does not contain format-4 error records for animals with an unknown sire.

Internet access to cow evaluations

A new application has been developed to allow a DRPC to obtain cow evaluations in format 105N from the AIPL web site (**http://aipl.arsusda.gov**) for cows in herds that are acquired after an evaluation cutoff. The original application to access cow evaluations processed a file of cow ID numbers that a requester uploaded to the web site. Several DRPC’s indicated that they would like to request records on a herd basis, and the new application allows selection of cows by herd code. Both applications return a file of format-105N records for downloading.

Registry status

With the use of international ID, an eartag ID number will be formatted the same as a registered ID number. The field for registry status in format 4T should be used to indicate

if the animal has been registered with a breed association. If the animal is known to be unregistered or the animal's registry status is unknown, the registry-status field should be left blank. Format-4T records with blank registry status are checked against pedigree files received from the breed associations. If a breed association has reported registry information for the animal, that registry status is used. Otherwise, the animal remains a grade until registered pedigree data are received. The coding schemes defined by the breed associations should be used when reporting registry status for registered animals.

Registration number limits

The current AIPL editing system uses limits on registration numbers to validate animal ID numbers, and these limits are updated whenever a breed association provides a file with records for new animals. With the new American ID, the limits on registration numbers are less effective because the range of numbers varies greatly depending on an organization's assignment of ID. So no limits are planned.

Payment for records in progress

The current system for counting the number of records in progress (RIP's) to determine National Association of Animal Breeders financial support to the DRPC's for providing RIP's requires that the RIP have a calving date within 15 months of the start of evaluation processing, 40 days or more in milk, and be usable for genetic evaluations. However, the current system is not able to distinguish records from owner-sampler herds that have not been included in genetic evaluations because of failure to meet herd edits for percentage of milk shipped or percentage of animals completely identified. The new editing system allows more accurate accounting of RIP's, and the following criteria are suggested for determining RIP's suitable for payment. Calving date plus days in milk must occur after the cutoff date for previous evaluation processing, the cow must be less than 96 months old at calving, and the RIP must be marked usable and have 40 days or more in milk and a termination code of 0. Using August 1997 evaluations as a test, the new accounting method resulted in an average decrease of 3.5% in the number of RIP's suitable for payment. Some of this difference can be attributed to unusable owner-sampler RIP's that had not been excluded by the current method.